COMP571/BIOC571 (Fall 2019)

Important note: The information contained in the course syllabus, other than the absence policies, may be subject to change with reasonable advance notice, as deemed appropriate by the instructor.

Who

Instructor:

- Huw A. Ogilvie
- Duncan Hall 3098
- hao3@rice.edu

TA:

- Zhen Cao
- Duncan Hall 3061
- zc36@rice.edu

Where and when

Distribution of class materials and assignment submission will be conducted via Canvas.

Seminars will be held in Duncan Hall **1046**, on Tuesdays and Thursdays, between 2:30–3:45 PM.

One scheduled office hour will be held each week, at 1:00 PM on Thursdays in Duncan Hall 3061. Individual appointments outside this time are welcome.

Intended audience

The students who should take COMP571/BIOC571 are generally studying computer science, biology or genomics, and wish to learn how to apply algorithms and statistical models to important problems in biology and genomics.

Course objectives and learning outcomes

The primary objective of the course is to teach the theory behind methods in biological sequence analysis, including sequence alignment, sequence motifs, and phylogenetic tree reconstruction. By the end of the course, students are expected to understand and be able to write basic implementations of the algorithms which power those methods.

Course materials

The main material for this course will be the course blog. But, if you wish to purchase a textbook, I highly recommend **Bioinformatics Algorithms** by Compeau & Pevzner. This is the recommended text for COMP416 “Genome-Scale Algorithms”, and I will give references to relevant chapters in the third edition.

Software for the course

Algorithms and statistics will be demonstrated using Python. Don’t worry if you are not fluent in either language, as no programs will have to be written from scratch.

The NumPy library for scientific computing will be used with Python. To install NumPy, first install the latest official distribution of Python 3. This can be downloaded for macOS or for Windows from Python.org, and should already be included with your operating system if you are using Linux.

Then simply use the Python package manage pip to install NumPy from the command line, by running `pip3 install numpy`.

Schedule

The course is organized around four themes;

1. Models and algorithms used for sequence alignment
2. Hidden Markov Models in computational biology
3. Phylogenetic trees and clustering
4. Optimality criteria and tree search methods

Each theme will have a corresponding homework assignment. Themes 1 and 2 will be covered in the first midterm, and themes 3 and 4 in the second midterm.

The below schedule may change subject to Rice University policy

Week	Tuesday class	Thursday class	Homework
08/26/2019	No class	Introduction, genomes, central dogma and homology	
09/02/2019	Empirical substitution matrices[1]	Global alignment[1]	
09/09/2019	Local alignment[1]	BLAST and BLAT[1]	
09/16/2019	PSSMs[1]	Pseudocounts[1]	#1 issued
09/23/2019	Hidden markov models[2]	Viterbi algorithm[2]	
09/30/2019	Forward algorithm[2]	Backward algorithm[2]	#1 due
10/07/2019	Applications of HMMs[2]	Midterm review[1,2]	#2 issued
10/14/2019	Midterm recess	Midterm exam[1,2]	
10/21/2019	Phylogenetic trees[3]	Guest lecture[3]	#2 due
10/28/2019	GTR substitution matrices[3]	Genetic distances[3]	
11/04/2019	Clustering (neighbor-joining)[3]	Clustering (UPGMA)[3]	#3 issued

Week	Tuesday class	Thursday class	Homework
11/11/2019	Parsimony algorithms[4]	The Felsenstein Zone[4]	
11/18/2019	Phylogenetic likelihood[4]	Hill climbing[4]	#3 due #4 issued
11/25/2019	MCMC[4]	Thanksgiving recess	
12/02/2019	Course review[3,4]	Final exam[3,4]	#4 due

Superscript numbers refer to the theme(s) for that day's class or midterm. Assignments will be both issued and due before midnight on Sundays.

Grade policies

- First in-class midterm: 25%
- Second in-class midterm: 25%
- Four homework assignments: 12.5% each

Students with a strong and valid excuse for not attending a midterm will be allowed to pick from one of the following options:

- Sit the midterm on a different day or time
- Adjust their grading to increase the contribution of the corresponding homework assignments to match the midterm's contribution
- Adjust their grading to double the contribution of the alternate midterm

Students with a strong and valid excuse for being unable to submit a homework assignment will be allowed to pick from one of the following options:

- Submit the homework assignment on a later day and time
- Adjust their grading to increase the contribution of the other homework assignments to match the assignment contribution
- Adjust their grading to increase the contribution of the corresponding midterm to match the assignment contribution

For both assignments and midterms the strength and validity of excuses, and which of the above options are made available, will be solely the instructor's purview. Without a strong and valid excuse, a penalty of 10 percentage points per day (which is equivalent to 1.25 points off the final course percent per day) will be applied to any assignment submitted after the deadline.

Absence policies

Attendance is expected at every class. Attendance for the midterm exams is compulsory and, without a strong and valid excuse, required to pass the course even if a student would have otherwise received a passing grade.

Rice Honor Code

In this course, all students will be held to the standards of the Rice Honor Code, a code that you pledged to honor when you matriculated at this institution. If you are unfamiliar with the details of this code and how it is

administered, you should consult the Honor System Handbook at http://honor.rice.edu/honor-system-handbook/. This handbook outlines the University's expectations for the integrity of your academic work, the procedures for resolving alleged violations of those expectations, and the rights and responsibilities of students and faculty members throughout the process.

Students with a disability

If you have a documented disability or other condition that may affect academic performance you should: 1) make sure this documentation is on file with Disability Support Services (Allen Center, Room 111 / adarice@rice.edu / x5841) to determine the accommodations you need; and 2) talk with me to discuss your accommodation needs.